Introduction to Data Science

June, 2017

About Me

- Eduction
  - 2012 Pass out, M.Sc. Information system - Bits, Pilani Rajasthan.
  - Trained in RHEL 6, AIX, Business Communications
  - Certified Data Modelling Engineer.
- Software Engineer
  - 4.5 Years in Data Engineering & Data Analytic.
  - 1 Year in Data Sciences and Data Modelling.
  - Python, Oracle DB, Oracle Apex.
- Personal Life
  - Teaching(blog), Music, Anime, lazy.
  - Health Conscious, Gym/Yoga/lots of Sleep.
  - Technology & Personal communication skills.
- Motivation:
  - Bridge the gap between Technology and People. Lead a R&D Team.

What is Data Science?

What is Data Science?

sci·ence (sī'əns)

1(a) The observation, identification, description, experimental investigation, and theoretical explanation of phenomena.

    Ex. New advances in science and technology.

 (b) Such activities restricted to a class of natural phenomena.

    Ex. The science of astronomy.

2. A systematic method or body of knowledge in a given area.

    Ex. The science of marketing.

3. Archaic Knowledge, especially that gained through experience.

Data Science Examples

Why Data Science?

Information Explosion & Doubling Processing Power

Metcalfe's law states that the value of a telecommunications network is proportional to the square of the number of connected users of the system (n2).

Moore's law is the observation that the number of transistors in a dense integrated circuit doubles approximately every two years.

Sources: Wikipedia

How to do Data Science

You can use lots of sophisticated analytical & Business Intelligent tools and come to a simple understandable explanations.

(or)

You can also use, simple tools like calculators or excel sheet to generate simple and simple results.

Data Science - In Action

Battles behind the scenes

Stages of Data Science

- Purpose
- Relevant Data Collection

- Reportings
- Finalising Report

- Wrangling(cleansing)*
- Data Analytics
- Feature Engg.*
- Data Modelling*
- Data Prediction*
- Evaluation*

- Data Product Building
  (software development)
  - Architecture
  - Development
  - Testing
  - Deployment

(*) ⇒ Repetitive stages

Titanic Example - Goal

Titanic Example – Data Preparation

Titanic Example – Data Modelling & Evaluation

Titanic Example – SCIKIT Example

Data Science - Tools

Too many to name, but none of them are close perfection.

Suggestions

Challenges in DS & Tips to who want to start.

Suggestions?

- Data Preparation
  - **"Give me six hours to chop down a tree and I will spend the first four sharpening the axe"**. Abraham Lincoln
  - Python, Scala, Excel, Databases(regex).
- Data Analytics
  - **"Seeing is believing"**
  - Python(Matplotlib, Seaborn), D3.Js, Excel.
- Data Models
  - **"There are no perfect solutions, but some work better"**
  - Learn 2-3 types of Clustering, Regression Models(LR,RF,SVM,KNN,XGB)
- Evaluation
  - **"A product not tested is broken by default"**
  - Accuracy, RMSE, Precision-Recall, F1 Score

Questions?

Sampath - Desk 4F 072. Imaginea Labs - Data Sciences.

Sachin, Keerat, Bipul, Kavi, Mageshwaran.

Thank you

Deep Learning Session on the way.